

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA
TERHADAP ISU RESESI TAHUN 2023 PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN *NAÏVE BAYES* CLASSIFIER**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SATRIA WAHYU RAHARJO

19.11.2614

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA
TERHADAP ISU RESESI TAHUN 2023 PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN *NAÏVE BAYES CLASSIFIER***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

SATRIA WAHYU RAHARJO

19.11.2614

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP
ISU RESESI TAHUN 2023 PADA MEDIA SOSIAL TWITTER
MENGUNAKAN *NAÏVE BAYES* CLASSIFIER**

yang disusun dan diajukan oleh

Satria Wahyu Raharjo

19.11.2614

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Agustus 2023

Dosen Pembimbing,



Norhikmah, M.Kom
NIK. 190302245

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP
ISU RESESI TAHUN 2023 PADA MEDIA SOSIAL TWITTER
MENGUNAKAN *NAÏVE BAYES CLASSIFIER*

yang disusun dan diajukan oleh

Satria Wahyu Raharjo

19.11.2614

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dina Maulina, M.Kom
NIK. 190302250

Ike Verawati, M.Kom
NIK. 190302237

Norhikmah, M.Kom
NIK. 190302245



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Satria Wahyu Raharjo
NIM : 19.11.2614

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Opini Masyarakat Indonesia Terhadap Isu Resesi Tahun 2023 Pada Media Sosial Twitter Menggunakan *Naive Bayes Classifier*

Dosen Pembimbing : Norhikmah, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 21 Agustus 2023

Yang Menyatakan,


Satria Wahyu Raharjo

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin, segala puji bagi ALLAH SWT berkat rahmat serta hidayah-Nya sehingga penyusunan skripsi ini dapat terselesaikan dengan lancar. Skripsi ini saya persembahkan untuk:

1. Kedua orang tua saya, selalu memberikan doa, restu, dan motivasi yang tak pernah berhenti kepada saya.
2. Ibu Norhikmah, M. Kom, yang menjadi dosen pembimbing saya, selalu dengan kesabaran dan ikhlas membimbing saya. Memberikan bantuan saat saya menghadapi kesulitan dalam menyusun skripsi, yang memungkinkan saya untuk menyelesaikan pekerjaan skripsi. Saya sangat berterima kasih atas ilmu yang telah dia bagikan kepada saya.
3. Rekan-rekan sekelas saya di 19-IF01, saya ingin mengucapkan terima kasih atas pertemanan dan keakraban yang telah terjalin selama 4 tahun ini. Bantuan serta dukungan yang diberikan oleh kalian sangat berarti bagi saya selama perkuliahan.
4. Para teman-teman Takeshi Castle yang selalu memberikan dukungan dan bantuan selama masa kuliah dan proses penyusunan skripsi. Saya sangat berterima kasih atas semua kenangan, persahabatan, dan ikatan kekeluargaan yang telah kita jalani selama ini.

KATA PENGANTAR

segala puji dan syukur penulis panjatkan kepada Allah SWT yang telah memberikan nikmat, rahmat, hidayah sehingga penulis dapat menyelesaikan skripsi ini. Penulis juga mengucapkan terimakasih kepada pihak-pihak yang telah membimbing dan membantu sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Isu Resesi Tahun 2023 Pada Media Sosial Twitter Menggunakan *Naïve Bayes Classifier*” dengan sebaik-baiknya. Penulis mengucapkan terimakasih atas segala dukungan dan bantuan yang diberikan selama mengerjakan penelitian ini kepada:

1. Bapak Prof. Dr. M. Suyanto, MM. selaku rektor Universitas Amikom Yogyakarta.
2. Ibu Norhikmah, M. Kom. selaku dosen pembimbing yang selalu membimbing penulis dalam melakukan penelitian ini.
3. Kedua orang tua saya, selalu memberikan doa, restu, dan motivasi yang tak pernah berhenti kepada saya.
4. Bapak dan Ibu dosen yang telah memberikan ilmu, pengetahuan, dan wawasan yang sangat berharga bagi penulis selama masa kuliah.
5. Semua pihak yang memberikan bantuan, dukungan, serta doa dalam proses penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini memiliki banyak kekurangan dan belum sempurna. Penulis mengharapkan kritik dan saran yang membangun untuk perbaikan laporan-laporan berikutnya.

Yogyakarta, 19 Agustus 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	12
2.2.1 Analisis Sentimen	12
2.2.2 <i>Text Mining</i>	12
2.2.3 <i>Preprocessing</i>	12
2.2.4 Resesi	13
2.2.5 Media Sosial.....	14

2.2.6	Twitter.....	14
2.2.7	Klasifikasi	14
2.2.8	<i>Naïve Bayes</i>	14
2.2.9	Python	16
2.2.10	<i>Confusion Matrix</i>	16
2.2.11	TF-IDF	16
2.2.12	VADER.....	17
2.2.13	SMOTE.....	18
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian.....	19
3.2	Alur Penelitian	19
3.2.1	Pengumpulan Dataset.....	20
3.2.2	Exploratory Data Analysis (EDA).....	20
3.2.3	Data <i>Preprocessing</i>	20
3.2.4	Translate Data	21
3.2.5	Labeling Otomatis.....	22
3.2.6	Pemodelan data	22
3.2.7	Evaluasi.....	27
3.3	Alat dan Bahan.....	28
3.3.1	Data penelitian	28
3.3.2	Perangkat Keras	28
3.3.3	Perangkat Lunak	28
BAB IV HASIL DAN PEMBAHASAN		29
4.1	Pengumpulan Dataset	29
4.2	Dataset	31
4.3	Exploratory Data Analysis (EDA).....	32
4.4	Pre processing data	35
4.4.1	Case Folding	35
4.4.2	Cleansing Data	36

4.4.3	Cek Data Kosong dan Data Duplikat.....	39
4.4.4	Tokenizing	40
4.4.5	Normalisasi	41
4.4.6	Stopword	44
4.4.7	Stemming	45
4.4.8	Cek Data Duplikat.....	48
4.5	Translate.....	49
4.6	Labeling Data.....	52
4.7	TF-IDF.....	54
4.8	Split Dataset.....	55
4.9	SMOTE.....	55
4.10	Klasifikasi <i>Naïve Bayes</i>	56
4.11	Evaluasi.....	57
4.12.1	Pengujian 90:10	57
4.12.2	Pengujian 80:20	58
4.12.3	Pengujian 70:30	60
BAB V PENUTUP		62
5.1	Kesimpulan	62
5.2	Saran	62
REFERENSI		64
LAMPIRAN.....		70

DAFTAR TABEL

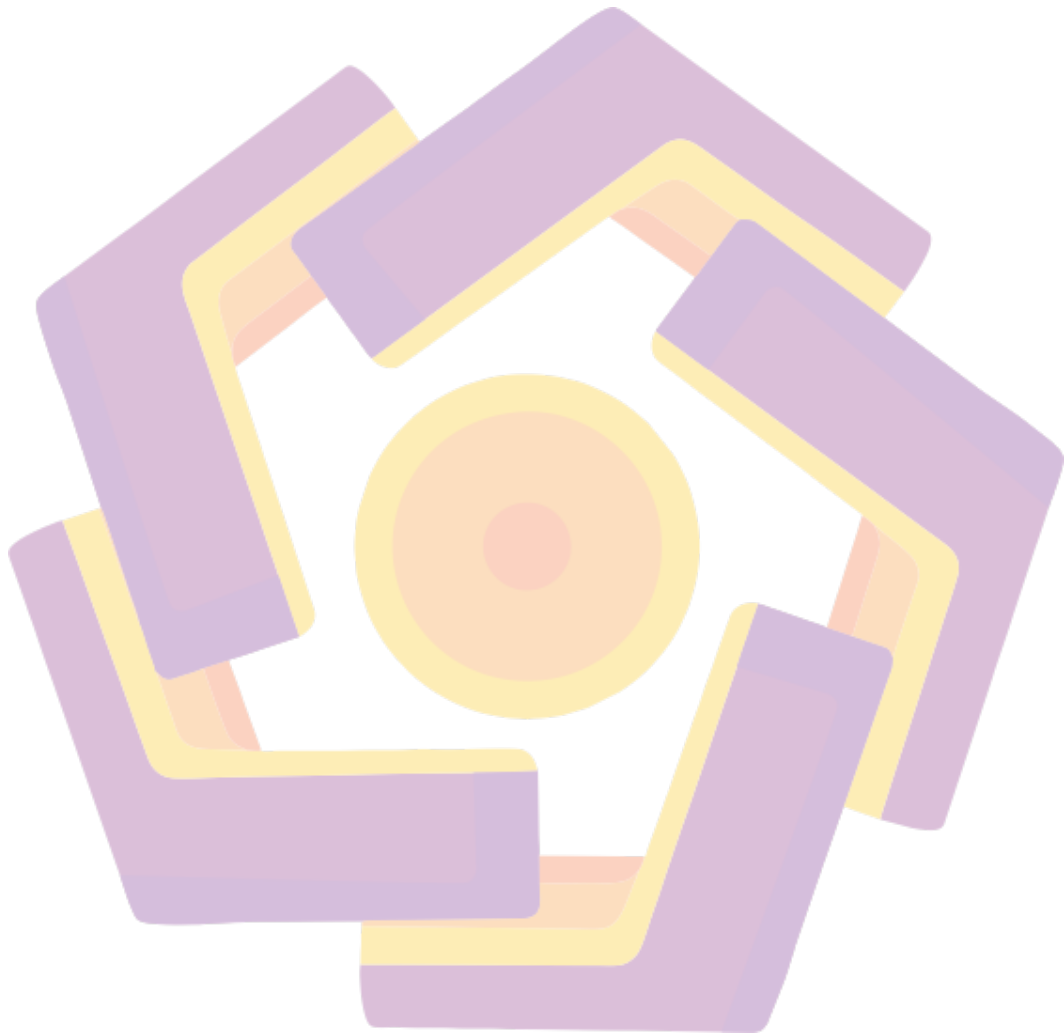
Tabel 2.1 Keaslian Penelitian	8
Tabel 2.2 Pembobotan Sentimen VADER [37]	18
Tabel 3.1 Hasil preprocessing	22
Tabel 3.2 Term Frequency (TF) dan Document Frequency (DF).....	22
Tabel 3.3 Perhitungan Inverse Document Frequency (IDF).....	23
Tabel 3.4 Perhitungan bobot term atau kata	24
Tabel 3.5 Tabel Data latih.....	25
Tabel 4.1 Dataset	32
Tabel 4.2 Hasil case folding.....	35
Tabel 4.3 Hasil cleansing data	38
Tabel 4.4 Hasil tokenizing	41
Tabel 4.5 Daftar kata new_kamusalay.csv.....	42
Tabel 4.6 Hasil proses normalisasi	43
Tabel 4.7 Hasil proses stopwords	45
Tabel 4.8 Hasil proses stemming	47
Tabel 4.9 Hasil proses terjemahan data	49
Tabel 4.10 Hasil proses pembersihan data terjemah.....	51

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	19
Gambar 4.1 Sumber kode install snsrape	29
Gambar 4.2 Sumber kode import.....	29
Gambar 4.3 Sumber kode crawling.....	30
Gambar 4.4 Sumber kode menampilkan data	30
Gambar 4.5 Sumber kode menyimpan csv	31
Gambar 4. 6 Dataset.....	31
Gambar 4. 7 Sumber kode EDA info dan shape	33
Gambar 4.8 Sumber kode EDA menghapus kolom.....	33
Gambar 4.9 Sumber kode EDA cek data duplikat dan isna.....	34
Gambar 4.10 Sumber kode EDA hapus data duplikat	34
Gambar 4.11 Sumber kode case folding	35
Gambar 4.12 Sumber kode hapus tweet khusus	36
Gambar 4.13 Sumber kode hapus nomor.....	36
Gambar 4.14 Sumber kode hapus tanda baca	36
Gambar 4.15 Sumber kode hapus spasi	37
Gambar 4.16 Sumber kode hapus dobel spasi	37
Gambar 4.17 Sumber kode hapus satu huruf	37
Gambar 4.18 Sumber kode penerapan semua fungsi def.....	38
Gambar 4.19 Sumber kode cek duplikat.....	39
Gambar 4.20 Sumber kode hapus duplikat	39
Gambar 4.21 Sumber kode mengubah spasi kosong menjadi nan.....	40
Gambar 4.22 Sumber kode cek nilai NaN	40
Gambar 4.23 Sumber kode hapus nilai NaN.....	40
Gambar 4.24 Sumber kode tokenizing.....	41
Gambar 4.25 Sumber kode baca new_kamusalay.csv	42
Gambar 4.26 Sumber kode proses normalisasi	42
Gambar 4.27 Sumber kode proses stopwords	44
Gambar 4.28 Sumber kode install.....	46

Gambar 4.29 Sumber kode stemming.....	46
Gambar 4.30 Sumber kode cek duplikat.....	48
Gambar 4.31 Sumber kode hapus data duplikat	48
Gambar 4.32 Sumber kode terjemah data.....	49
Gambar 4.33 Sumber kode pembersihan tanda	50
Gambar 4.34 Sumber kode pembersihan duplikat.....	51
Gambar 4.35 Sumber kode import library	52
Gambar 4.36 Sumber kode hitung polarity.....	52
Gambar 4.37 Sumber kode proses labeling	53
Gambar 4.38 Sumber kode hasil VADER.....	53
Gambar 4.39 Sumber kode merubah tipe data.....	54
Gambar 4.40 Sumber kode proses TF-IDF.....	54
Gambar 4.41 Sumber kode splitting 80:20	55
Gambar 4.42 Sumber kode SMOTE.....	56
Gambar 4.43 Sumber kode <i>Naïve Bayes</i>	56
Gambar 4.44 Hasil pengujian 90:10 tanpa smote	57
Gambar 4.45 Hasil pengujian 90:10 menggunakan smote	58
Gambar 4.46 Hasil pengujian 80:20 tanpa smote	59
Gambar 4.47 Hasil pengujian 80:20 menggunakan smote	59
Gambar 4.48 Hasil pengujian 70:30 tanpa smote.....	60
Gambar 4.49 Hasil pengujian 70:30 menggunakan smote	61

DAFTAR LAMPIRAN



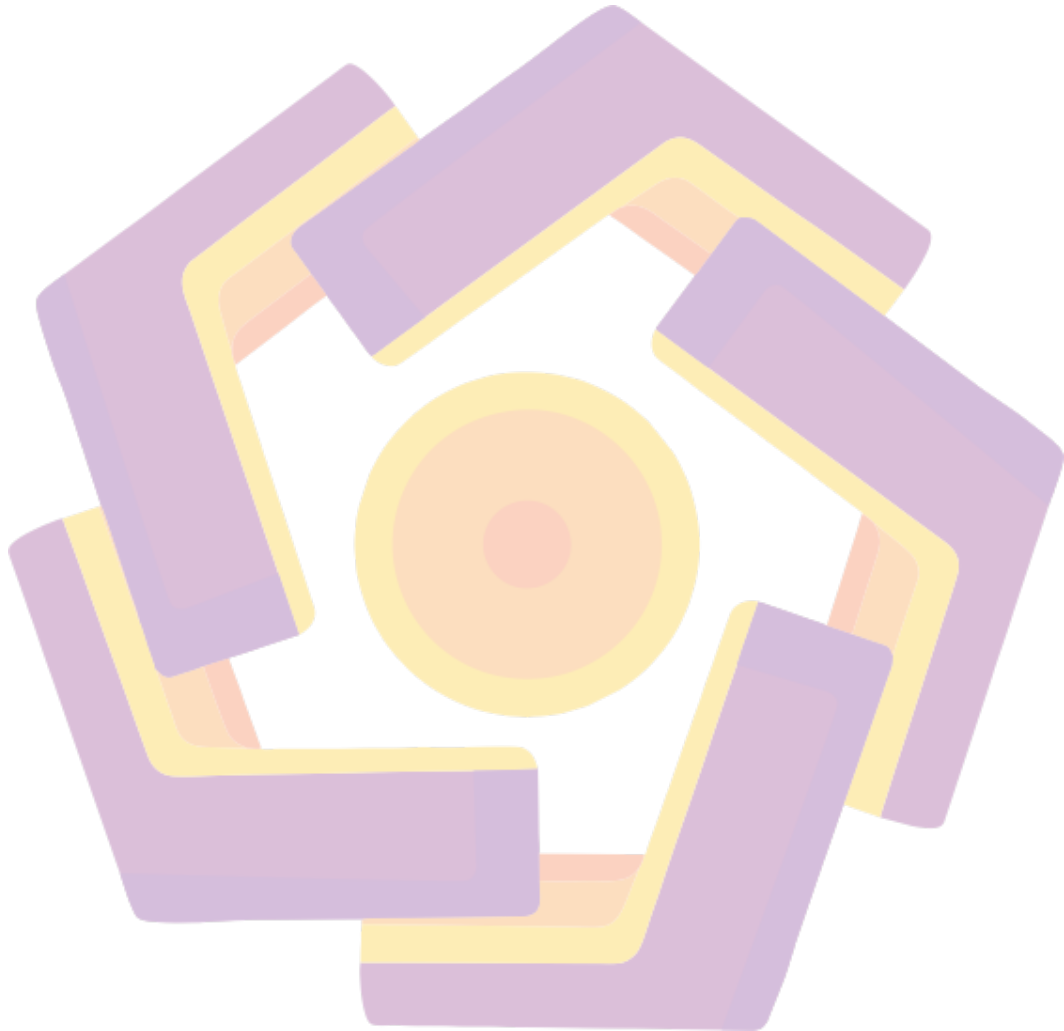
DAFTAR LAMBANG DAN SINGKATAN

sklearn	scikit-learn
SMOTE	Synthetic Minority Oversampling Technique
VADER	Valence Aware Dictionary for Sentiment Reasoning
TF-IDF	Term Frequency-Inverse Document Frequency
EDA	Exploratory Data Analysis
NLTK	Natural Language Toolkit



DAFTAR ISTILAH

Tweet	Komentar pada media sosial twitter
Preprocessing	Demprosesan data



INTISARI

Perkembangan teknologi dan informasi yang semakin pesat seperti sekarang ini, masyarakat dapat dengan mudah memberikan opini, kritikan maupun masukan suatu isu yang sedang hangat atau ramai diperbincangkan melalui media sosial. Twitter menjadi salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia. Salah satu yang sempat ramai di media sosial twitter adalah isu resesi 2023. Banyaknya opini dari pengguna twitter mengenai topik isu resesi 2023 dapat dilakukan penelitian yang tujuannya untuk analisis sentimen mengenai isu resesi 2023 di media sosial twitter Indonesia.

Penelitian analisis sentimen ini menggunakan metode *Naïve Bayes* yaitu salah satu metode yang digunakan untuk klasifikasi berdasarkan dari teorema bayes yang semua kegiatan memberikan kontribusi yang sama penting pada pemilihan target kelas tertentu. *Naïve Bayes* menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes. Dataset yang digunakan diambil dari media sosial twitter dengan kata kunci resesi 2023 dengan cara *crawling* data. Jumlah dataset yang digunakan sebanyak 3853 data mentah atau kotor. Hasil dari beberapa tahap yang dilakukan pada penelitian ini sehingga dataset label Negative berjumlah 2208 data, label Positive berjumlah 577 data, dan label Neutral berjumlah 215 data. Performa paling bagus ketika menggunakan algoritma *Naïve Bayes* dipadukan dengan SMOTE terjadi pada pengujian 80:20 dengan akurasi sebesar 0.816, precision 0.494, recall 0.89, dan f1_score 0.63.

Kata kunci: analisis sentimen, *Naïve Bayes*, twitter, resesi, teks mining.

ABSTRACT

The rapid development of technology and information as it is today allows people to easily provide opinions, criticisms, or feedback on a hot or widely discussed issue through social media. Twitter is one of the social media platforms widely used by the Indonesian community. One of the topics that gained attention on Twitter was the 2023 recession issue. The abundance of opinions from Twitter users regarding the recession issue provides an opportunity for research aimed at sentiment analysis regarding the 2023 recession issue on Indonesian Twitter.

This sentiment analysis research utilizes the Naïve Bayes method, which is one of the techniques used for classification based on Bayes' theorem where all activities contribute equally to the selection of a specific target class. Naïve Bayes employs probability and statistics calculations as proposed by Thomas Bayes. The dataset is obtained from Twitter with the keyword "resesi 2023" (recession 2023) using data crawling. The total number of raw or uncleaned datasets used is 3853. Through several stages in this research, the dataset is labeled, resulting in 2208 data labeled as Negative, 577 data as Positive, and 215 data as Neutral. The best performance is achieved when using the Naïve Bayes algorithm combined with SMOTE in the 80:20 testing scenario, yielding an accuracy of 0.816, precision of 0.494, recall of 0.89, and an f1_score of 0.63.

Keyword: sentiment analysis, Naïve Bayes, Twitter, recession, text mining.